

Feedback Control of a Chemical Mechanical Polishing Device

Providing Manipulation of Removal Rate Profiles

Inventors: Arulkumar P. Shanmugasundram, Alexander T. Schwarm,
Gopalakrisna B. Prabhu

Related Applications

10 This application is a continuation-in-part application of and claims priority under 35 U.S.C. § 119(e) from provisional application serial no. 60/298,878 filed June 19, 2001.

Field of the Invention

15 The present invention generally relates to the planarization of substrates, in particular, semiconductor wafers, and more particularly to a method and apparatus for providing feedback control of the planarization process.

Background of the Invention

Chemical-mechanical polishing (CMP) is used in semiconductor fabrication
20 processes for obtaining full planarization of a semiconductor wafer. The method involves removing material, e.g., a sacrificial layer of surface material, from the wafer (typically, silicon dioxide (SiO_2)) using mechanical contact and chemical erosion. Polishing flattens out height differences, since areas of high topography (hills) are removed faster than areas of low topography (valleys).

CMP typically utilizes an abrasive slurry dispersed in an alkaline or acidic solution to
25 planarize the surface of the wafer through a combination of mechanical and chemical action. Generally, a CMP tool includes a polishing device (having an attached wafer to be polished) positioned above a rotatable circular platen on which a polishing pad is mounted. In use, the platen may be rotated and an abrasive slurry is introduced onto the polishing pad. Once the
30 slurry has been applied to the polishing pad, a downward force may be applied to a rotating

head to press the attached wafer against the pad. As the wafer is pressed against the polishing pad, the wafer is mechanically and chemically polished.

The effectiveness of a CMP process may be measured by its polishing rate, and by the resulting finish (absence of small-scale roughness) and flatness (absence of large-scale topography) of the substrate surface. The polishing rate, finish and flatness are determined by a variety of factors, including the pad and slurry combination, the relative speed between the substrate and pad and the force pressing the substrate against the polishing pad.

As semiconductor processes are scaled down, the importance of CMP to the fabrication process increases. In particular, it is increasingly important to control and minimize within wafer (WIW) thickness non-uniformity. A variety of factors may contribute to producing variation across the surface of a wafer during polishing. For example, variations in the surface topography may be attributed to drift of the processing conditions in the CMP device. Typically, the CMP device is optimized for a particular process, but because of chemical and mechanical changes to the process, e.g., changes in the polishing pad during polishing, degradation of process consumables, and other factors, the CMP process may drift from its optimized state. In addition to processing drift, the wafer surface coming into the CMP process may be non-uniform, which exacerbates the process-induced variations across the post-polished surface.

Recent attempts to correct processing drift include feedback control, in which information generated during current processing is used to adjust future processing runs. One control variable used in such feedback control of the polishing step includes the arm oscillation length of the polishing tool. Feedback loops have also been developed for optimization of polishing pad conditioning. However, these schemes are still not adequate in today's manufacturing environment to satisfactorily compensate for the aforementioned effects.

The present invention is directed to overcoming, or at least reducing the effects of, one or more of the problems set forth above.

Summary of the Invention

The present invention relates to a method, apparatus and medium for planarizing a surface of a substrate, for example, a semiconductor wafer, in order to improve run-to-run control over the wafer thickness profile. The present invention uses a model (which can be 5 implemented as a single model or multiple models) of the planarization process to predict material removal across the wafer surface and to improve within wafer thickness uniformity. Deviations from the predicted outcome are used to set new polishing parameters, which feedback into the process to enhance process results.

In one aspect of the present invention, a method of producing a uniform wafer 10 thickness profile in a polishing operation includes (a) providing a model for a wafer polishing that defines a plurality of regions on a wafer and identifies a wafer material removal rate in a polishing step for each of the regions, and (b) polishing a wafer using a polishing recipe that generates a target thickness profile for each region.

In another aspect of the present invention, a method of controlling surface non-uniformity of a wafer in a polishing operation includes (a) providing a model for a wafer 15 polishing that defines a plurality of regions on a wafer and identifies a wafer material removal rate in a polishing step of a polishing process for each of the regions, wherein the polishing process includes a plurality of polishing steps, (b) polishing a wafer using a first polishing recipe based upon an incoming wafer thickness profile, (c) determining a wafer 20 thickness profile for the post-polished wafer of step (b), and (d) calculating an updated polishing recipe based upon the wafer thickness profile of step (c) and the model of step (a) to maintain a target wafer thickness profile.

In at least some embodiments of the present invention, the first polishing recipe is 25 based on the model of step (a) to obtain the target wafer thickness profile, or the first polishing recipe is determined empirically.

In at least some embodiments of the present invention, the plurality of regions in the model of step (a) includes regions extending radially outward from a center point on the wafer. The model may include four or more regions.

In at least some embodiments of the present invention, the polishing of step (b) includes polishing the wafer at a plurality of polishing stations. The polishing step may be carried out at three polishing stations.

5 In at least some embodiments of the present invention, the polishing recipe is the same at at least two polishing stations.

In at least some embodiments of the present invention, the polishing recipe is different at at least two polishing stations.

10 In at least some embodiments of the present invention, calculating the updated polishing recipe of step (c) includes calculating updated polishing recipes for each of the plurality of polishing stations.

In at least some embodiments of the present invention, the updated polishing recipes for each of the plurality of polishing stations accounts for the tool state of the individual polishing stations. The wafer thickness profile for each of the subsequent polishing stations may be provided by the prediction from previous stations.

15 In at least some embodiments of the present invention, the step of providing a model includes (e) measuring pre-polished wafer thickness in each of a plurality of regions defined on one or more wafers, (f) polishing the one or more wafers, wherein polishing includes polishing the one or more wafers in a plurality of polishing steps, (g) measuring the wafer material removal rate for the one or more wafers at each of the plurality of regions after each 20 of the polishing steps of step (g), (h) providing a model defining the effect of tool state on polishing effectiveness, and (i) recording the pre-polished and post-polished wafer thicknesses for each of the regions on a recordable medium. The model may further include fitting the data to a linear or non-linear curve that establishes a relationship between the material removal rate of a region of the wafer and a polishing parameter of interest.

25 In at least some embodiments of the present invention, polishing parameter includes polishing time. The polishing parameters may further include a parameter selected from the group consisting of polishing time, polishing pad down forces and velocity, slurry flow and composition, conditioning time, conditioning disk down forces and velocity, oscillating speeds of both the conditioning disk and the wafer carrier.

In at least some embodiments of the present invention, wafer removal for a region j (AR'_j) in the model of step (a) is determined according to the equation:

$$AR'_j = (c_{11j} \cdot x_1 + c_{12j}) \cdot t_1 + (c_{21j} \cdot x_2 + c_{22j}) \cdot t_2 + (c_{31j} \cdot x_3 + c_{32j}) \cdot t_3 + (c_{41j} \cdot x_4 + c_{42j}) \cdot t_4 + (c_{51j} \cdot x_5 + c_{52j}) \cdot t_5,$$

where x_1, x_2, x_3, x_4 , and x_5 are the additional parameter values for polishing steps 1, 2, 5, 3, 4, and 5, respectively; t_1, t_2, t_3, t_4 , and t_5 are the polishing times for polishing steps 1, 2, 3, 4, and 5, respectively, and c_{aj} provides the contribution to wafer removal of the variable x in polishing step a in region j ; and c_{a2j} provides the contribution to wafer removal of polishing time in polishing step a . The wafer material removal rate profile may account for tool state by scaling the profile using the scaling factor:

$$10 \quad (1 + k_p \cdot t_p + k_d \cdot t_d + k_{pd} \cdot t_p \cdot t_d),$$

where the terms t_p and t_d refer to pad and disk life, respectively, with units of hour; and the terms k_p , k_d and k_{pd} are empirically determined coefficients relating pad and disk life to removal rate.

15 In at least some embodiments of the present invention, an updated polishing recipe is attained by solving the equation:

$$\min_x f(y^{sp}, g(x))$$

where x is a vector of times and other processing parameters corresponding to the polishing recipe; $g(x)$ is the model for the polishing process, y^{sp} is a vector of the desired average region wafer thicknesses; and $f(y^{sp}, g(x))$ is a penalty function to penalize the deviation 20 between the model predictions $g(x)$ and the desired thicknesses y^{sp} .

In another aspect of the present invention, a method of determining a model for wafer thickness profile includes (a) measuring pre-polished wafer thickness in each of a plurality of regions defined on one or more wafers, (b) polishing the one or more wafers, wherein polishing includes polishing the one or more wafers in a plurality of polishing steps, (c) 25 measuring the wafer material removal rate for the one or more wafers at each of the plurality of regions after each of the polishing steps of step (b), (d) providing a model defining the effect of tool state on polishing effectiveness, and (e) recording the pre-polished and post-polished wafer thicknesses for each of the regions on a recordable medium. The model may

include fitting the data to a linear or non-linear curve that establishes a relationship between the material removal rate of a region of the wafer and a polishing parameter of interest.

In at least some embodiments of the present invention, the polishing parameter includes polishing time. The polishing parameters may include a parameter selected from 5 the group consisting of polishing time, polishing pad down forces and velocity, slurry flow and composition, conditioning time, conditioning disk down forces and velocity, oscillating speeds of both the conditioning disk and the wafer carrier.

In at least some embodiments of the present invention, the wafer material removal for a region j (AR'_j) in the model of step (a) is determined according to the equation:

10
$$AR'_j = (c_{11j} \cdot x_1 + c_{12j}) \cdot t_1 + (c_{21j} \cdot x_2 + c_{22j}) \cdot t_2 + (c_{31j} \cdot x_3 + c_{32j}) \cdot t_3 + (c_{41j} \cdot x_4 + c_{42j}) \cdot t_4 + (c_{51j} \cdot x_5 + c_{52j}) \cdot t_5,$$

where x_1 , x_2 , x_3 , x_4 , and x_5 are the additional parameter values for polishing steps 1, 2, 3, 4, and 5, respectively; t_1 , t_2 , t_3 , t_4 , and t_5 are the polishing times for polishing steps 1, 2, 3, 4, and 5, respectively, and c_{aj} provides the contribution to wafer removal of the variable x in polishing step a in region j ; and c_{a2j} provides the contribution to wafer removal of polishing time in polishing step a . The wafer material removal rate profile may account for tool state 15 by scaling the profile using the scaling factor:

$$(1 + k_p \cdot t_p + k_d \cdot t_d + k_{pd} \cdot t_p \cdot t_d),$$

where the terms t_p and t_d refer to pad and disk life, respectively, with units of hour; and the terms k_p , k_d and k_{pd} are empirically determined coefficients relating pad and disk life 20 to removal rate.

In at least some embodiments of the present invention, a the model is determined using less than 10 wafers.

In one aspect of the present invention, an apparatus for conditioning polishing pads used to planarize substrates is provided having a carrier assembly having a plurality of arms 25 for holding a wafer positionable over a plurality of planarizing surfaces of a plurality of polishing pads, controlling means capable of controlling an operating parameter of the polishing process, and a controller operatively coupled to the controlling means, the controller operating the controlling means to adjust the operating parameter of the polishing

process as a function of a model for a wafer thickness profile, the model including defining a polishing model that defines a plurality of regions on a wafer and identifies a wafer material removal rate in a polishing step of a polishing process for each of the regions, wherein the polishing process includes a plurality of polishing steps.

5 In at least some embodiments of the present invention, the model defines wafer removal for a region j (AR'_j) in the wafer material removal rate model according to the equation:

$$AR'_j = (c_{11j} \cdot x_1 + c_{12j}) \cdot t_1 + (c_{21j} \cdot x_2 + c_{22j}) \cdot t_2 + (c_{31j} \cdot x_3 + c_{32j}) \cdot t_3 + (c_{41j} \cdot x_4 + c_{42j}) \cdot t_4 + (c_{51j} \cdot x_5 + c_{52j}) \cdot t_5,$$

10 where x_1, x_2, x_3, x_4 , and x_5 are the additional parameter values for polishing steps 1, 2, 3, 4, and 5, respectively; t_1, t_2, t_3, t_4 , and t_5 are the polishing times for polishing steps 1, 2, 3, 4, and 5, respectively, and c_{aj} provides the contribution to wafer removal of the variable x in polishing step a in region j ; and c_{a2j} provides the contribution to wafer removal of polishing time in polishing step a .

15 In another aspect of the present invention, a computer readable medium includes instructions being executed by a computer, the instructions including a computer-implemented software application for a chemical mechanical polishing process, and the instructions for implementing the process include (a) receiving data from a chemical mechanical polishing tool relating to the wafer removal rate of at least one wafer processed in the chemical mechanical polishing process, and (b) calculating, from the data of step (a), 20 updated polishing recipe, wherein the updated polishing recipe is calculated by determining the difference between an output of a wafer material removal rate model and the data of step (a).

25 In at least some embodiments of the present invention, the model for a wafer material removal rate defines a plurality of regions on a wafer and identifies a wafer material removal rate in a polishing step of a polishing process for each of the regions, wherein the polishing process includes a plurality of polishing steps.

In at least some embodiments of the present invention, the wafer removal for a region j (AR'_j) in the wafer material removal rate model is determined according to the equation:

$$AR'_j = (c_{11j} \cdot x_1 + c_{12j}) \cdot t_1 + (c_{21j} \cdot x_2 + c_{22j}) \cdot t_2 + (c_{31j} \cdot x_3 + c_{32j}) \cdot t_3 + (c_{41j} \cdot x_4 + c_{42j}) \cdot t_4 + (c_{51j} \cdot x_5 + c_{52j}) \cdot t_5,$$

where x_1 , x_2 , x_3 , x_4 , and x_5 are the additional parameter values for polishing steps 1, 2, 3, 4, and 5, respectively; t_1 , t_2 , t_3 , t_4 , and t_5 are the polishing times for polishing steps 1, 2, 3, 4, and 5, respectively, and c_{alj} provides the contribution to wafer removal of the variable x in polishing step a in region j ; and c_{a2j} provides the contribution to wafer removal of polishing time in polishing step a .

The term “target wafer thickness profile” represents the desired processing outcome of the CMP process. Some tolerance is built into the profile, so that a feedback control system defines a target profile and acceptable standard deviations therefrom, wherein such deviations would not require updating of the polishing recipe. Use of the term target wafer thickness profile includes the target and the standard deviation therefrom.

The term wafer is used in a general sense to include any substantially planar object that is subject to polishing. Wafers include, in addition to monolith structures, substrates having one or more layers or thin films deposited thereon. Throughout the specification, wafer and thin film may be used interchangeably, unless otherwise indicated.

“Tool state” refers to the condition of the consumable or variable components of the CMP apparatus. Most often this term is used to refer to the state of the conditioning disk and polishing pad, which change continually over the lifetime of the pads, and idle time. Typical conditioning disk life is about 60 hours and typical polishing pad life is about 30 hours.

20

Brief Description of the Drawings

Various objects, features, and advantages of the present invention can be more fully appreciated with reference to the following detailed description of the invention when considered in connection with the following drawing, in which like reference numerals identify like elements. The following drawings are for the purpose of illustration only and 25 are not intended to be limiting of the invention, the scope of which is set forth in the claims that follow.

FIG. 1 is a perspective view of a chemical mechanical polishing apparatus.

FIG. 2 is a plot of oxide material removal (\AA) across the surface of a substrate for successive polishing steps in a polishing recipe.

FIG. 3 is a flow diagram generally illustrating model development;

FIG. 4 is a schematic illustration of a wafer showing regions defined for thickness profile model.

FIG. 5 is a flow diagram of the feedback loop used in a CMP polishing operation, as
5 contemplated by at least some embodiments of the present invention.

FIG. 6 is a schematic illustration of model development for a CMP process using two platens with different polishing recipes, as contemplated by at least some embodiments of the present invention.

FIG. 7 is a block diagram of a computer system that includes tool representation and
10 access control for use in at least some embodiments of the invention.

FIG. 8 is an illustration of a floppy disk that may store various portions of the software according to at least some embodiments of the invention.

Detailed Description of the Invention

15 FIG. 1 shows a perspective view of a typical CMP apparatus **100** for polishing one or more substrates **110**. The CMP apparatus **100** includes a series of polishing stations **101** and a transfer station **102** for loading and unloading substrates. Each polishing station includes a rotatable platen **103** on which is placed a polishing pad **104**. A source of polishing fluid **111** may be provided to supply polishing fluid **112** to the polishing pad **104**. Each polishing
20 station may include an associated pad conditioning apparatus **105** to maintain the abrasive condition of the polishing pad. A rotatable multi-head carousel **106** is supported by center post **107** about which the carousel rotates. The carousel **106** includes multiple carrier heads **108** that are capable of independently rotating about its own axis. The carrier head **108** receives a substrate from and delivers a substrate to the transfer station **102**. The carrier head provides a controllable load, i.e., pressure on the substrate to push it against the polishing pad when the polishing station and the carrier head are engaged. Some carrier heads include a retaining ring **109** to hold the substrate and help to provide the polishing load. To effectuate polishing, the platen **103** may be rotated (typically at a constant speed). Moreover, individually variable down forces may be applied by each of the carrier heads **108**, for
25

example by adjusting retaining ring pressures. The carrier heads **108** holding substrates **110** can rotate on axis **113** and oscillate back and forth in slot **114**.

One type of CMP process polishes the wafer in a series of polishing steps. By way of example, FIG. 2 shows a CMP profile for eight successive polishing steps **201** through **208** for a single wafer 200 mm in diameter. Each polishing step removes a subset of the total material to be polished from the substrate surface. Moreover, the thickness profile generated by each polishing step may be different, as is seen by comparison of profiles **201** and **208**. The final, thin film thickness profile is the sum of the individual polishing step thickness profiles and desirably produces a uniform wafer thickness across the surface.

A CMP process may include the transport of a sample from polishing station (platen) to polishing station (platen). One type of CMP process distributes wafer removal among the various platens, and each platen will have a full set of polishing steps to achieve the desired material removal for that platen. Any combination of removal is possible. Thus, by way of example, where it is desired to remove 6000 Å of material in total, 3000 Å may removed from the polishing station at platen 1, 1000 Å may be removed at platen 2, and 2000 Å may be removed at platen 3. The polishing recipe for each platen may be the same or different.

The CMP processes described above may be modeled to provide a format for improving the planarization process. The model can be represented as raw data that reflects the system, or it can be represented by equations, for example multiple input-multiple output linear, quadratic and non-linear equations, which describe the relationship among the variables of the system. By using a model, the within wafer thickness uniformity can be improved or maintained run-to-run by adjusting the polishing parameters during wafer polishing to correct for unmodeled effects or to correct for drift in the polishing process conditions. By way of example, polishing time, polishing pad down forces and velocity, slurry flow and composition, conditioning time, conditioning disk down forces and velocity, oscillating speeds of both the conditioning disk and the wafer carrier may be adjusted during the polishing operation in a feedback and feedforward loop that predicts and then optimizes the polishing recipe.

According to at least some embodiments of the present invention, an initial model is developed based upon knowledge of the wafer polishing process, as is shown in a flow

diagram (FIG. 3). An initial understanding of the system is acquired in step **300**, which is used to design and run a design of experiments (DOE) of step **310**. The DOE desirably is designed to establish the relationship between or among variables that have a strong and predictable impact on the processing output one wishes to control, e.g., wafer thickness. The
5 DOE provides data relating to process parameters and process outcome, which is then loaded to the advanced process control system in step **320**. The advanced processing control system may be a controller or computer that uses the data to create and update the model.
Processing requirements such as output targets and process specification are determined by the user in step **325**, which are combined with the DOE data to generate a working model in
10 step **330**.

An illustrative example of model development is now described. According to at least some embodiments of the present invention, a model structure is defined that models wafer material removal rate (polishing) profiles as independent steps in the CMP process. As described herein above (FIG. 2), the individual steps may be combined to produce a uniform, final wafer thickness. The steps to be used in the model can also be defined as subsets of removal rate profiles; that is, a step may consist of a family of removal rate profiles that have similar characteristics. For each family of removal rate profiles, polishing parameters are identified, which may be varied, and their effect on the outcome is determined. Exemplary polishing variables, which may be included in this model include, but are not limited to,
15 polishing time, polishing pad down forces and velocity, slurry flow and composition, conditioning time, conditioning disk down forces and velocity, and sweep speeds of both the conditioning disk and the wafer carrier.
20

In at least some embodiments of the present invention, the model relies on removal rate profiles based on regions of the wafer. As is shown in FIG. 4, a wafer may be divided into radial regions **401** through **405** of varying width and area. The number of regions is not set for the model and may be selected based upon the polishing profile. Thus, for example, FIG. 2 designates seven (7) regions across the wafer, while FIG. 4 illustrates five (5). The size and location of the regions also may vary and may be selected based upon the effect of certain polishing parameters on the wafer in that region.

The number, size and location of regions may be selected based upon the complexity of the wafer material removal rate profile. In at least some embodiments, it is desirable that the profile in any given region be substantially uniform, particularly in those cases where a number of wafer thickness measurement within a region are averaged to define the region-
5 averaged thickness profile. Thus, at the edges where edge effects can be dramatic, narrow regions encompassing only the outer regions may be selected. Near the center of the wafer where polishing effects may be more subtle, a larger region may be defined. The regions are defined such that all azimuthal variation is averaged out since the CMP tool can not correct for such variation. Film thickness measurements taken within a region of the wafer are
10 averaged to give the average thickness for that region.

To obtain DOE data, a polishing step is run and, based upon incoming measurements, e.g., pre-polishing and post-polishing wafer thickness measurements, and processing parameter values, a removal rate profile or, equivalently, a wafer thickness profile, can be determined for each region. Conventionally, data may be acquired empirically, by carrying out a series of experiments over a range of parameter values and over the lifetime of the polishing pad and conditioning disk. Such an approach makes no assumptions about the processing characteristics of the polishing operation, and the data is fit to the appropriate curve to define the model. This approach requires a large number of wafers, at least 30 for a 4 step process, and is time consuming (a typical disk life is about 60 hours).

In at least some embodiments of the present invention, a modified approach to obtaining DOE data is used. The approach assumes that the data may be fit to a linear curve and that superposition is valid. Superposition assumes that the same results are attained by performing a first step for a set time, followed by performing a second step for a set time, e.g., separately, but sequentially, as are attained by running the two steps together. In
25 addition, the approach uses an established model to relate pad and disk life to polishing performance. These assumptions significantly reduce the amount of data (and hence number of samples) required to model the system appropriately. In at least some embodiments of the present invention, it is sufficient to run less than 10, and even 6-8 wafer for proper model development. By way of example only, the DOE may include 5-7 polishing steps and the
30 polishing recipe may be carried out on a few wafers, as few as one, or for example 5-8 wafers. More wafers are required for polishing recipes with more polishing steps.

By way of example, a series of experiments may be conducted for a polishing system of interest as described above for determining the relationship between wafer material removal rate profile and polishing time and other parameters of interest. Standard polishing procedures may be used, with all polishing pad and wafer conditions held constant with the exception of the parameter(s) under investigation. Exemplary polishing parameters that may be held constant include polishing pad size, polishing pad composition, wafer composition, pad conditioning time, rotational velocity of the polishing pad, and rotational velocity of the wafer. In at least some embodiments of the present invention, the parameters under investigation include at least the polishing time for each of the polishing steps in the polishing recipe and the polishing down force (P), as defined by retaining ring pressure. As is shown in greater detail in the analysis that follows, additional parameters may be incorporated into the model.

Once data from DOE runs are obtained, a model may be developed. A model having five-polishing steps may be defined as follows:

$$AR'_j = c_{1j} \cdot t_1 + c_{2j} \cdot t_2 + c_{3j} \cdot t_3 + c_{4j} \cdot t_4 + c_{5j} \cdot t_5, \quad (1)$$

Where AR'_j is the amount of material removed for region j of the wafer; t_1, t_2, t_3, t_4 , and t_5 are the polishing times for polishing steps 1, 2, 3, 4, and 5, respectively; and $c_{1j}, c_{2j}, c_{3j}, c_{4j}$, and c_{5j} are removal rates for region j in polishing steps 1, 2, 3, 4, and 5, respectively.

Additional parameters may be included in the model, and the model may be defined as follows:

$$AR'_j = (c_{11j} \cdot x_1 + c_{12j}) \cdot t_1 + (c_{21j} \cdot x_2 + c_{22j}) \cdot t_2 + (c_{31j} \cdot x_3 + c_{32j}) \cdot t_3 + (c_{41j} \cdot x_4 + c_{42j}) \cdot t_4 + (c_{51j} \cdot x_5 + c_{52j}) \cdot t_5, \quad (2)$$

where x_1, x_2, x_3, x_4 , and x_5 are the additional parameter values for polishing steps 1, 2, 3, 4, and 5, respectively; t_1, t_2, t_3, t_4 , and t_5 are the polishing times for polishing steps 1, 2, 3, 4, and 5, respectively, and c_{aj} provides the contribution to wafer removal rate of the variable x in polishing step a in region j ; and c_{a2j} provides the contribution to wafer removal rate of polishing time in polishing step a . Thus, the model permits inclusion of an unlimited number of processing parameters.

In at least some embodiments of the present invention, the model may be further augmented to include the effect of the tool state. The tool state represents the effect of wear,

use and aging on the tool, and includes the condition of the conditioning disk and polishing pad, represented by disk life and pad life, and also includes idle time. This functionality may be expressed as a scaling factor. An exemplary scaling factor that takes pad life and disk life into account is shown in eq. 3.

5
$$AR_j = (1 + k_p \cdot t_p + k_d \cdot t_d + k_{pd} \cdot t_p \cdot t_d) \cdot (AR'_j |_{t_p=0, t_d=0}) \quad (3)$$

The terms t_p and t_d refer to pad and disk life, respectively, with units of hour; and the terms k_p , k_d and k_{pd} are empirically determined coefficients relating pad and disk life to removal rate, or equivalently, to the amount of material removed. AR'_j at $t_p=0$ and $t_d=0$ is the amount of material removed under initial polishing conditions.

10 Using a model such as the one just described, a prediction for region-averaged, final wafer thickness can be calculated given incoming thickness, the pad and disk life, the polishing step times and the value for any other selected parameter for those steps which vary that parameter.

15 Process model development and optimization are carried out with reference to a specific polishing system. That is, the conditions that effect within wafer uniformity are specific to the type of wafer being polished, the slurry used in polishing and the composition of the polishing pad. Once a wafer/slurry/polishing pad system is identified, the system is characterized using the models developed according to the invention. In at least some embodiments of the present invention, it is envisioned that a separate model (or at least a 20 supplement to a composite model) is created for each slurry/polishing pad wafer combination (i.e., for each different type/brand of slurry and each type/brand of pad that may be used in production with a given type of wafer).

25 Also, at least some embodiments of the present invention contemplate a wafer polishing model that can accommodate polishing at multiple platens, either in parallel or serially. The CMP process often consists of multiple platens, which are operated simultaneously. Typically, each platen removes a portion of the total amount of material to be removed. The wafers are advanced from platen to platen, and each platen has a separate recipe that determines the polishing step times and other processing parameters, such as retaining ring pressures for each of the steps that are performed on that platen.

A process model that accounts for the effects of multiple platens that perform similar or different polishing steps on wafer thickness profile is illustrated in FIG. 6. In a first phase 600 of the model, the polishing recipe 610 (here, 6 steps) for platen 1 620 is determined (the “first polishing process”). Process input data 630, such as incoming wafer thickness for the defined regions of the pre-polished wafer, disk life and pad life, are input into the model. 5 The wafer is polished and final wafer thicknesses 640 for each of the wafer regions is measured. Post-polished regions thicknesses 640 from the first polishing process are used as input data in a second phase 645 of the model development. A second polishing recipe 650 is carried out on platen 2 660, which can be the same as or different from that carried out on 10 platen 1 620. Pad life and disk life factors 655 relating to the pad and conditioning disk used on platen 2 660 are also included in the model. Final thickness measurements 670 are taken and used in the model development. Thus, the method of the invention can accommodate a model that involves multiple polishing processes on multiple platens having different tool states and is able to provide platen-specific feedback 680 and 690 to platens 1 and 2, 15 respectively. The model is extremely versatile and able to accommodate highly complex polishing scenarios.

According to at least some embodiments of the present invention, an initial model developed as described herein above is used in at least some embodiments of the present invention to control the run-to-run uniformity of the polishing process and to provide a 20 feedback loop for updating the polishing recipe (FIG. 5). Briefly, one or more wafers is processed according to a first polishing recipe. A thickness measurement is taken across the polished wafer to obtain a wafer thickness profile, which is compared to the predicted wafer thickness profile calculated by the model. If the measured wafer thickness profile indicates deviation from the desired results, those deviations are used in an optimization process to 25 update the polishing recipe. The updated recipe is then used in a feedback loop to progressively optimize the polishing recipe so as to improve or maintain within wafer film thickness uniformity.

According to the processing flow diagram in FIG. 5, initial processing conditions, e.g., tool state and wafer state are identified that will provide a desired wafer removal rate 30 profile in step 500. The initial conditions may be determined empirically or by using the processing model of at least one embodiment of the present invention. If a processing model

is used, a controller can use this model to calculate step times and processing parameters to polish an incoming profile to a target flat profile with a desired thickness as shown in step 510. Wafers are polished according to the initial polishing recipe in the CMP tool at step 520. The thickness of the polished wafer is measured and deviation from the predicted thickness is determined in step 530. In step 540 it is determined whether the deviation exceeds an established tolerance. If the deviation is within acceptable ranges, no changes are made to the polishing recipe and the controller instructs the tool to reuse the existing recipe in step 550. If the deviation is outside acceptable limits, new target parameters are set in step 560 and are feedback in step 570 into the controller where the polishing recipe is optimized 10 according to an updated model that takes the deviation from the predicted value into consideration. The polishing step may be repeated and further updates of the polishing recipe are possible.

Process control of the CMP process according to at least one embodiment of the present invention permits optimization of the wafer removal rate for series of regions j across the wafer surface. By individually optimizing for the regions j of the wafer, greater control over the total surface is attainable. Thus, greater within wafer uniformity is achieved.

An exemplary optimization method, which can be used in determining an updated model for determining an updated polishing recipe, solves the equation:

$$\min_x f(y^{sp}, g(x)) \quad (4)$$

where x is a vector of times and other processing parameters corresponding to the polishing recipe; $g(x)$ is the model for the CMP process as described above in eqs. 1-3; y^{sp} is a vector of the desired average region wafer thicknesses; and $f(y^{sp}, g(x))$ is some function which is meant to penalize the deviation between the model predictions $g(x)$ and the desired thicknesses y^{sp} .

Thus, the optimization method suggests that the model need not correct for 100% of the deviation from predicted value. A penalty function may be used to reflect uncertainty in the measured or calculated parameters, or to “damp” the effect of changing parameters too quickly or to too great an extent. It is possible, for example, for the model to overcompensate for the measured deviations thereby necessitating another adjustment to

react to the overcompensation. This leads to an optimization process that is jumpy and takes several iterations before the optimized conditions are realized.

Based upon this optimization method, the post-polishing wafer thickness is measured and the difference between the predicted thickness and the final thickness is determined. The
5 error in prediction, also known as a bias, is then linearly added into the model such that the predicted final thickness more closely matches the actual final thickness. This bias is added to each region j , which is modeled as is shown in the following equation:

$$FT_j = IT_j - AR_j + b_j \quad (5)$$

where FT_j is the predicted final thickness of region j ; IT_j is the incoming thickness of region
10 j ; AR_j is the predicted amount which is removed from region j given a set of recipe parameters; and b_j is the bias term which arises due to the difference between the predicted and actual amount removed from region j . The process of linearly updating a model with bias terms based upon the difference between a model prediction and an actual measurement is part of at least some feedback controls in at least some embodiments of the present
15 invention.

In at least some embodiment of the present invention, a feedback control combines the platens into a single model using the average of the tool states for each of the platens. The single model would use the feedback approach described above to apportion the bias adjustment across the different platens in some predetermined way.

20 Also, in at least some embodiments of the present invention, a feedback control scheme uses the final thickness measurements to distribute feedback individually to all of the platens. A method for modeling a polishing process for which different platens are separately modeled and factored into the model is shown in FIG. 6. Because each platen can be treated individually, the tool state, e.g., the pad life and disk life, and idle time, can
25 be included in the model and feedback can be specific to the platen and polishing recipe. This feedback control scheme is particularly useful when different polishing recipes are being carried out on each platen. The ability to separately model each platen provides a greater degree of processing flexibility, since it allows one to change the processing recipe at one platen (during one stage of the polishing process) while keeping the processing recipe at
30 the remaining platens unchanged.

In at least some embodiments of the present invention, the controller applies feedback individually to each carrier head. Each carrier head performs in a unique manner and it is possible in updating the polishing recipe to separately review the past performance of each wafer carrier and to adjust the updated parameters accordingly.

5 Feedback and feedforward control algorithms are constructed for use in the above control process based on the above models using various methods. The algorithms may be used to optimize parameters using various methods, such as recursive parameter estimation. Recursive parameter estimation is used in situations such as these, where it is desirable to model on line at the same time as the input-output data is received. Recursive parameter
10 estimation is well suited for making decisions on line, such as adaptive control or adaptive predictions. For more details about the algorithms and theories of identification, see Ljung
L., *System Identification - Theory for the User*, Prentice Hall, Upper Saddle River, N.J. 2nd edition, 1999.

15 In at least some embodiments of the present invention, the polishing recipe may be updated in discrete increments or steps defined in the algorithms of the developed by the model. Also, in at least some embodiments of the present invention, the updated polishing recipes may be determined by interpolation to the appropriate parameters.

20 Additional apparatus utilized to implement the feedforward and feedback loop include a film thickness measurement (metrology) tool to provide thickness data needed to calculate wafer material removal rate. The tool may be positioned on the polishing apparatus so as to provide in-line, *in situ* measurements, or it may be located remote from the polishing apparatus. The tool may use optical, electrical, acoustic or mechanical measurement methods. A suitable thickness measurement device is available from Nanometrics (Milpitas, CA) or Nova Measuring Instruments (Phoenix, AZ). A computer may be utilized to calculate
25 the optimal pad conditioning recipe based upon the measured film thickness and calculated removal rate, employing the models and algorithm provided according to the invention. A suitable integrated controller and polishing apparatus (Mirra with iAPC or Mirra Mesa with iAPC) is available from Applied Materials, California.

Exemplary semiconductor wafers that can be polished using the concepts discussed herein including, but are not limited to those made of silicon, tungsten, aluminum, copper, BPSG, USG, thermal oxide, silicon-related films, and low k dielectrics and mixtures thereof.

The invention may be practiced using any number of different types of conventional
5 CMP polishing pads. There are numerous polishing pads in the art that are generally made of urethane or other polymers. Exemplary polishing pads include Epic™ polishing pads (Cabot Microelectronics Corporation, Aurora IL) and Rodel® IC1000, IC1010, IC1400 polishing pads (Rodel Corporation, Newark, DE), OXP series polishing pads (Sycamore Pad), Thomas West Pad 711, 813, 815, 815-Ultra, 817, 826, 828, 828-E1 (Thomas West).

10 Furthermore, any number of different types of slurry can be used in conjunction with aspects of the present invention. There are numerous CMP polishing slurries in the art, which are generally made to polish specific types of metals in semiconductor wafers. Exemplary slurries include Semi-Sperse® (available as Semi-Sperse® 12, Semi-Sperse® 25, Semi-Sperse® D7000, Semi-Sperse® D7100, Semi-Sperse® D7300, Semi-Sperse® P1000, 15 Semi-Sperse® W2000, and Semi-Sperse® W2585) (Cabot Microelectronics Corporation, Aurora IL), Rodel ILD1300, Klebesol series, Elexsol , MSW1500, MSW2000 series, CUS series and PTS (Rodel).

Various aspects of the present invention that can be controlled by a computer can be (and/or be controlled by) any number of control/computer entities, including the one shown
20 in FIG. 7. Referring to FIG. 7 a bus 756 serves as the main information highway interconnecting the other components of system 711. CPU 758 is the central processing unit of the system, performing calculations and logic operations required to execute the processes of embodiments of the present invention as well as other programs. Read only memory (ROM) 760 and random access memory (RAM) 762 constitute the main memory of the
25 system. Disk controller 764 interfaces one or more disk drives to the system bus 756. These disk drives are, for example, floppy disk drives 770, or CD ROM or DVD (digital video disks) drives 766, or internal or external hard drives 768. These various disk drives and disk controllers are optional devices.

A display interface 772 interfaces display 748 and permits information from the bus
30 756 to be displayed on display 748. Display 748 can be used in displaying a graphical user

interface. Communications with external devices such as the other components of the system described above can occur utilizing, for example, communication port **774**. Optical fibers and/or electrical cables and/or conductors and/or optical communication (e.g., infrared, and the like) and/or wireless communication (e.g., radio frequency (RF), and the like) can be used as the transport medium between the external devices and communication port **774**.

5 Peripheral interface **754** interfaces the keyboard **750** and mouse **752**, permitting input data to be transmitted to bus **756**. In addition to these components, system **711** also optionally includes an infrared transmitter and/or infrared receiver. Infrared transmitters are optionally utilized when the computer system is used in conjunction with one or more of the processing

10 components/stations that transmits/receives data via infrared signal transmission. Instead of utilizing an infrared transmitter or infrared receiver, the computer system may also optionally use a low power radio transmitter **780** and/or a low power radio receiver **782**. The low power radio transmitter transmits the signal for reception by components of the production process, and receives signals from the components via the low power radio receiver. The low power

15 radio transmitter and/or receiver are standard devices in industry.

Although system **711** in FIG. 7 is illustrated having a single processor, a single hard disk drive and a single local memory, system **711** is optionally suitably equipped with any multitude or combination of processors or storage devices. For example, system **711** may be replaced by, or combined with, any suitable processing system operative in accordance with the principles of embodiments of the present invention, including sophisticated calculators, and hand-held, laptop/notebook, mini, mainframe and super computers, as well as processing system network combinations of the same.

FIG. 8 is an illustration of an exemplary computer readable memory medium **884** utilizable for storing computer readable code or instructions. As one example, medium **884** may be used with disk drives illustrated in FIG. 7. Typically, memory media such as floppy disks, or a CD ROM, or a digital video disk will contain, for example, a multi-byte locale for a single byte language and the program information for controlling the above system to enable the computer to perform the functions described herein. Alternatively, ROM **760** and/or RAM **762** illustrated in FIG. 7 can also be used to store the program information that is used to instruct the central processing unit **758** to perform the operations associated with the instant processes. Other examples of suitable computer readable media for storing

information include magnetic, electronic, or optical (including holographic) storage, some combination thereof, etc. In addition, at least some embodiments of the present invention contemplate that the medium can be in the form of a transmission (e.g., digital or propagated signals).

5 In general, it should be emphasized that various components of embodiments of the present invention can be implemented in hardware, software or a combination thereof. In such embodiments, the various components and steps would be implemented in hardware and/or software to perform the functions of the present invention. Any presently available or future developed computer software language and/or hardware components can be employed
10 in such embodiments of the present invention. For example, at least some of the functionality mentioned above could be implemented using the C, C++, or any assembly language appropriate in view of the processor(s) being used. It could also be written in an interpretive environment such as Java and transported to multiple destinations to various users.

15 Although various embodiments which incorporate the teachings of the present invention have been shown and described in detail herein, those skilled in the art can readily devise many other varied embodiments that incorporate these teachings.